


```

DB 61 SACEPCGALGATFDKLLFEYGEVETLAERAKAGSVVLGPIVNHGRGLNRCGFESFE 120
OY 121 DPELAGAALALIRIGIOSTGVOATIKHFLCNDQEDRRMVGSIYTERALREYALPFOIA 180
DB 121 DPELAGAALALIRIGIOSTGVOATIKHFLCNDQEDRRMVGSIYTERALREYALPFOIA 180
DB 121 DSTLSGLAASVYLIGLQSNKVGACIKHFVCMEDERNSSVSDVOSRALREYALPFOIA 180
OY 181 VRDSQGAFTAVNGINGVSCSENPXYLDGMLRKEMGMGLMSDMGYSTTEAVNAGL 240
DB 181 CKTSNFKSLMTSTNKVNGEHSOSRILDLNLRKEMMGSTIISDFGATSLKALDAGL 240
OY 241 DLEMPGPRFRG-ETLKFNVSNGKPRTHVIDQAREVLOFVKCAASGVTEGPEPTVNN 299
DB 241 DLEMPGPRFRNVTIQLHVGSKELSESILDERAKNVLLKVRHSMQNTAEHNCLELN-ND 299
OY 300 TPEPTAALLKRVNEGIVLAKNENNVPLSKKKKTLIVGNAAQATYHGGGSAALRYAV 359
DB 300 SSGLRALKKFAQSIVLLKNNKRLPLSRKGFVAVIGNAKVCNYSGGGSANLKRYYTV 359
OY 360 TPEPGLSKOLETPSYTGA--YTPVPLIGEOCLTPPDAPGMWFEVNEPPTPRKOH 417
DB 360 SMTDGLIAKIDGVEYALGCHNTLNPNTANLNVNRTGKHGVAKFYLEPATSEKRTLI 419
OY 418 DELFTKTDHLYDYHHPKAAATW-YADMEGYTADDECTYELGIVVCGTAKAAYVDOLV 476
DB 420 DDVLEDGVYRFYDYNCKMKDGYFIDEGYLIPDEDAVYERGISVFGTALLFIDVLL 479
OY 477 VDNATQVPGDAPFGSATRETRIRLVGNNTYKKEIEGSAPTYLLKQDTIVPGSGSR 536
DB 480 IDKTKQTPNHTFEETGTEERNSTYLRKGRKNVREYEGSAATYTLNPLSPSTGRYS 539
OY 537 VGSCKTVIDDOAEIEKSVLAKENDVOYIICAGLNADMETGADRASMKLGVLDOLADVA 596
DB 540 ICGVKTVIDPELIDYAVRAKASVDCYIICVGLFAEMETEGEDRKTYLPLSLDKLYVSL 599
OY 597 AANPNVYVWQSTPEEMWLDATPAVIAQWYNGNFTGNSIADVYEGDYNPSGKLSLSP 656
DB 600 QSNPNVYVWQSTPEEMWISAEHTLHMYNGNELGNALNIIFGEQNPCKGLPITFP 659
OY 657 KRLDNPAPLNFRTGAGRTLYGEDVYVGYRYEFAKDNFPGHGLSTTTFASLSLVS 716
DB 660 KRLDNPAPLNFRTGAGRTLYGEDVYVGYRYEFAKDNFPGHGLSTTTFASLSLVS 716
OY 717 HKDKLSVLSVKNGTGVSVAQVADLYVPLQAKINRPVKELGKPAVELOPGETKAVT 776
DB 720 NCEBRLRIDELISNTGMSGALIIQYISQI-VRSVNRPVKELKEPSKYLVCPEKTKLIR 778
OY 777 IEEDKTVAAVYFDEERDQCEKGYEVIVSSSAKADGVALRGKTYGETYMWSGV 833
DB 779 IEIDIKYVNSFYDELNEKWCSEGEYVNLVTS--KD-IALTGKFTLKTTHMTGL 832

```

RESULT 2

```

ID 09C329 PRELIMINARY: PRT: 862 AA.
AC 09C329;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Beta-glucosidase (EC 3.2.1.21).
OS Volvariella volvacea.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Pleurotaceae; Volvariella.
OX NCBI_TaxID=36659;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V14;
RA Ding S., Ge W., Buswell J.A.;
RT "Molecular cloning, characterization and expression of a beta-
RT glucosidase gene from the edible straw mushroom, Volvariella
RT volvacea."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF329731; AAG59831.1; -

```

```

DR Interpro: IPR002772; GH_3C.
DR Interpro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3_1.
DR Pfam: PF01915; Glyco_hydro_3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
DR PROSITE: PS00775; GLYCOSYL-HYDROL_F3; 1.
KW Glycosidase; Hydrolase.
SQ
SEQUENCE 862 AA; 94848 MW; BB3C4B08113C32PC CRC64;

Query Match 38.8%; Score 1703; DB 3; Length 865;
Best Local Similarity 42.0%; Pred. No. 3e-105;
Matches 372; Conservative 158; Mismatches 272; Indels 84; Gaps 20;

OY 2 ADIDEALIKKTLAEKVDLAGIDFMTKALPHGVPSLFTDGPNGVGRKFEKNCVPA 61
DB 7 AKANIDEVLEQTLIDEALSILGAGFMHTHAIELSGAPAVAVSDGPNIGNHFMGTPTA 66
OY 62 ACEPGTSLGSTFNQTLLEAG-KMKGEALAKSANHYLGFTIMQSPSLGGRFESIGE 120
DB 67 KLPSTSLGATWDPVEVVEEGLKLPAEAKLRASISLVAFSTNIQRNPLGSGSFESFE 126
OY 121 DPELAGAALALINGIOSTGVOATIKHFLCNDQEDRRMVGSIYTERALREYALPFOIA 180
DB 127 DPLSGIISASVYNGVQKGGIGATIKHFVGNDRKEDDROGDSIISERALREYALPMLT 186
OY 181 VRDSQGAFTAVNGINGVSCSENPXYLDGMLRKEMGMGLMSDMGYSTTEAVNAGL 240
DB 187 OKYAPAPAMINAVNNGVNHVAEDPELLQVLRNEMKYGKLSMDWGMVSDHGIAGL 246
OY 241 DLEMPGPRFRG-ETLKFNVSNGKPRTHVIDQAREVLOFVKCAASGVTE--NPEPT 295
DB 247 DLEMPGIMKMTLDDVNVNTIQRKLTDPDIRARVYLEVKKC-ACGAPETLIDDEER 305
OY 296 TYNNPETAALLRKVNGEIVLKNENNVPLSKK---KTLVGNPAQATYHGGGSA 351
DB 306 TV-ELESKLMRLKRASSIVLKNND-NVPLKBEGGAIKKIAYVGNKARQVLSGGGSA 363
OY 352 ALRAYAVTPEDGSLKOLE--TPPSYTVG--ATTVVPI-----LGEK----- 390
DB 364 ALKASYFISPDGIIKAALPEHGVETFESEGAARAKTLPTLEMDLETEGEGMIGTWH 423
OY 391 ---CLTDDGAPGMWRFNE-----PPGPNRHOIDELEFTTDMHLYDYHHPKA 437
DB 424 ESDDSMTLDDPEFLAPRLVDETRIFISTYKGTIKR-----WTNPLKGYLAPRE 473
OY 438 ADTWYADMEGYTADDECTYELGIVVCGTAKAAYVDOLVDNATQVPGDAPFGSATRE 497
DB 474 KDT-----NFEGLIAGRAKLMWDGOLVIDNMTRRORGEAFPGSGSQQE 518
OY 498 TGRINLVKGNNTYKKEIEFGS--APTYLLKQDTIVPGHGLRVGCGKYIIDQAEIEKVAL 555
DB 519 TGVYLLKAGKKHEIYVEKCNRAPADGDEDEALINDSNPGVRLGAEVAMADLLSEAVKL 578
OY 556 AKENHOVITICAGLNADMETEGADRASMKLPGVLDOLADVAANPNVYVWQSTPEEMP 615
DB 579 ASEADAVIAVGLNADMETEGADRTALPGRTBLEKVAKNSKVYVQAQSAITLP 638
OY 616 WIDATPAVIAQWYNGNFTGNSIADVYEGDYNPSGKLSLSPFKRLDNPAPLNFRTGAGRT 675
DB 639 WIDSAAYVHANVYLCGNATGDAIADYLPCKQNPCKLSLTPFKRLIEDVPNSHGFSENGKV 698
OY 676 LYGEDVYVGYRYEADKDNVFPFGHGLSYTTFASFNSLVS----HKDG--KLSVSLV 728
DB 699 RYAEDLFVGKYKHHRNIEPLFPFGFGLSYTTFSSLNQISAPVIDHATSSFSITATLSI 758
OY 729 KNTGVSVAQVADLYVPLQAKINRPVKELGKPAK-ELDPGETKAVTIEEDKTVAAVY 787
DB 759 TNGTGVTSSEVAOLVSVYPTSELTTHAPLQLRARKKVADLPGETREKTL-ELDKVAVSY 817
OY 788 FDEEDQCEKGYEVIVSSSAKADGVALRGKTYGETYMWSGV 833
DB 818 WNDRIQTAVENGELIVGNSSVSKD-LTVQRFVYKDGEMWGI 862

```

RESULT 3

09HGX2

ID 09HGX2 PRELIMINARY: PRT: 861 AA.

AC 09HGX2;

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, last annotation update)

DE Putative beta glucosidase (EC 3.2.1.21).

GN BGL.

OS Agaricus bisporus (Common mushroom).

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Agaricales; Agaricaceae; Agaricus.

OX NCBI_TaxID=5341;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D649;

RA Morales-Almora P., Thurston C.F.;

RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ293760; CAC03462.1;

DR InterPro: IPR002772; GH_3C.

DR InterPro: IPR001764; GH_3N.

DR Pfam: PF00933; Glyco_hydro_3; 1.

DR Pfam: PF01915; Glyco_hydro_3; C; 1.

DR PRINTS: PR00133; GLHYDRASE3.

DR Glycosidase; Hydrolase.

KW SEQUENCE 861 AA; 94353 MW; 8B52C0A18DB39AB3 CRC64;

Query Match

Best local similarity 36.3%; Score 1595.5; DB 3; Length 861;

Matches 365; Conservative 147; Mismatches 306; Indels 51; Gaps 22;

QY 2 ADDIVEAIKKLLTAEKVDLLAGIDFHTKALPKHGVPSLFTDGPNGVGTFFNGVPA 61
 DB 7 ANADIDTIVALTDEADILLISAGVGFHTAIEQLIPAKVSDGPRGNGHFMGTPTA 66
 QY 62 ACFPCGCTSLSTGNOTLLEBAG-KMMGKEAIKSAHVILGPTINMORSLGSGFESIGE 120
 DB 67 KCLPSSNMGATPRDLLEEVGLKILAREKLKLSASVITLPTONIQNPLGSGFESFE 126
 QY 121 DPLAGLGAALALRGISTGVQATIKHFLCNDQEDRRMVOISVTERALREIYALPQIA 180
 DB 127 DPLSGMLCSAYINGIKGIGATIKHFGVNDKEDRTGYDSI-HYXTSSGDLSPSVHA 185
 QY 181 VRDSQP--GAFMTAYNGINVSCESENKYLDMKLRKMGMDGLMSWYTYTTEVA 238
 DB 186 GSRNMPHGAIMTAAYNVNNGIHVCENPELLQKILIDEMKSDVMIMSDFGTYSVDGLNA 245
 QY 239 GLDLEMPGPPRERG-ETLKFNVNSNGKPFIVIDQAREVLOFQKCAASVTE---NGP 293
 DB 246 GLDLEMPGLNKRSLSEVNSISQSKRYTAKKVERAKVLELYKC-AQAPRLDGDGL 304
 QY 294 ETTVNNPTETAALLRKVNGEIVLLKNENNVPLSK--KTLIVGNKQATYHGGSA 351
 DB 305 ERTL-DSDEDNALRKFAESIYLLKNDKRYLLPDRKSLKVAIVGNNAKAAVLSGGSA 363
 QY 352 ALRAYAVTPEDGSKOL-ETPP-----SYTVGATTT-VPRILGEOCLTPRGAGM--RWR 403
 DB 364 ALKSTYVARPAGLVKALKEVNPVIEVYSEGAFAAKTLPLDPLETSEGGKGTGSMH 423
 QY 404 VFNP-----GTPNRQIHIDELFTKDMHLYDYHPRKADTWYADMEGTYADE-DCY 456
 DB 424 AHEMDDSLVALPTPIKQYVDE-----TRMFISSSPAGITTKKWTLMKQYILPREDC 478
 QY 457 YELGLVVGCAKAYVDOLVVDNAKQVPGDAFGSATREBTGRINLVKNTYKFKTEFG 516
 DB 479 FEFGLTAAGRAKLFTVDKLVDMNTRORGVEFFGCGSEETGVVELKAGVHEIYVDFC 538
 QY 517 S--APTTLKGDITVPGGHSRGGVGVVIDQAEIEKSVLAKHDDVYIICAGINAMWT 574
 DB 539 NVRGPADEDETETIMDSNPGIRLGALEVEDEDELLNSAVSTAKADAVIIVGLNGDMET 598

QY 575 EGADRAKMKLEGVLDOLIAADVAANPNVYMQGTPEEMRMIDATPAVIAQYGNETG 634
 DB 599 EGNDDTTLALPGRTELDVQKVAANPRAVVTQASATIMPVDDVSLHTLWLGNTG 658
 QY 635 NSIADVVGCDVNPISGKLSLSPKRLQDNPAFLNPETEAGRTLYGDEVYGRYFEADKD 694
 DB 659 DAIAIVIGGNHNPISGKLSLTPKRLLEDVPAHGFSEKGTAYAGLTVGKHIOHRNIE 718
 QY 695 -VNEPFGHGLSYTTFAFNSLVSHK-----DGKLSVLSVYNTSGVPAQVLAQLYKPL 747
 DB 719 PLFATGCHGLSTYSFINSIDRLSLQPLVSGDQPLKATVRLT--NIGNITGSOYVQLICLP 776
 QY 748 QAAKINRPYKELKGRPAK-ELQPGETKAVTIEQEKYAAVFEEDQMCYKIDYEVIV 806
 DB 777 KTSLETHPRWQLRGFKKMDVAPGESREVEL-VMDRLSVYWDKE---WVENGAIDVAV 832
 QY 807 SDSS--AAKQVALLGKFLYGETYMWSSV 833
 DB 833 APTSEGVGEGELLGRFVNEKFGWRL 861

RESULT 4

ID 066050 PRELIMINARY: PRT: 772 AA.

AC 066050;

DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, last annotation update)

DE Beta-glucosidase.

GN GLDA.

OS Ruminococcus albus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Lachnospiraceae; Ruminococcus.

OX NCBI_TaxID=1264;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=7;

RA Cho K.K.;

RT "Sequence and structure of Ruminococcus albus 7.";

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: U92808; AAC05445.1;

DR InterPro: IPR002772; GH_3C.

DR InterPro: IPR001764; GH_3N.

DR Pfam: PF00933; Glyco_hydro_3; 1.

DR Pfam: PF01915; Glyco_hydro_3; C; 1.

DR PRINTS: PR00133; GLHYDRASE3.

KW SEQUENCE 772 AA; 86092 MW; 0E6CD8BEF923CC3 CRC64;

Query Match

Best local similarity 26.0%; Score 1143.5; DB 2; Length 772;

Matches 261; Conservative 134; Mismatches 243; Indels 187; Gaps 14;

QY 4 IDEVALTKLLTAEKVDLLAGIDFHTKALPKHGVPSLFTDGPNGV---GTFKN-G 58
 DB 13 MDIAHMETMTLEKSLSCSGADFWHTKATIERLDIPQIVSOGPHLRKNVDSNDPNA 72
 QY 59 VPACFPCTSLSTGNOTLLEBAGKMMGKEAIKSAHVILGPTINMORSLGSGFESI 118
 DB 73 IEAVCFPTAALACSVDRELKIDKALGEGQSEKVSVILPGCVIKRSLPCGRFEYF 132
 QY 119 GEDPFLAGLGAALALRGISTGVQATIKHFLCNDQEDRRMVOISVTERALREIYALPQ 178
 DB 133 SEDPYLASOMALSHIRIGVOSKAGTSLKHPAANNQHRMSVSAEIDETHEIYLAPE 192
 QY 179 IAVRDSQPGAFMTAYNGINVSCESENKYLDMKLRKMGMDGLMSWYTYTTEVA 238
 DB 193 SVTKEKPMTWKSYNKGESISSQKSLITLTLRKMGFDGLVMSDKAVDDPKAGITA 252
 QY 239 GLDLEMPGPPRERG-ETLKFNVNSNGKPFIVIDQAREVLOFQKCAASVTEGPEYTVN 298
 DB 253 GLDLEMPGSKCKNDKMLKLAIVEDGKLSVEALDKCVRIELIDKSLDEC-----RTEM 305
 QY 299 NPET--AALLRKVNGEIVLLKNENNVPLSKKKTLLIYGNKQATYHGGSAALRAY 357

Db 306 WDKERHQAQKAERSAVLLKNDHILLPSKNEKIAFGARAEOPRYOGSSSHINSR 365
 QY 358 AVTPEDGLSKOLETPPSYTVGATTVTPILGECCLTPDAGPMRMVRVNEPGRTPNRQHI 417
 Db 366 TVSALE----- 371
 QY 418 DELEFTKTMHLVDYHPRADTWYADMEGYTADEDCYELGLVVCSTAKAYVDQLVY 477
 Db 372 -----AVDGM-----ENITYAKG----- 384
 QY 478 DNATQVPGDAFPGSATRETRERINLVKGNTRYKFIERSASATYTLKGDYIPGHSILRY 537
 Db 385 -----FSLNDEI----- 392
 QY 538 GGCKVIDDOAEIE-KSVALAKEHDVILICAGLNADMETEGADRASMKLPGLVDQLADVA 596
 Db 393 -----NTELEQCAVEAMNADKYVYFAGLPDSFESGCFRKMOLPOCQIDILDKLS 444
 QY 597 AANPTVVVMQGTPEBFWL-----DATPAVIOAMYGNETGNSIADVYFGDYVPSGK 650
 Db 445 EVNPNIVVVLHNGAPVEMFANGDEDSNVKAILMEYLSGQAAGBAVYRILLEGVNPSSGK 504
 QY 651 LSLSPFKRQDNPAPLNFTEAGRTLYGEDVYVGRYRYEFADKDVNFPFGHLSYTTAF 710
 Db 505 LAETPEPLRLENDPSTYLNFPGEADIVKYSGLIFVGRYREKKMVELYPPGHLSYTEFEY 564
 QY 711 SNLSVSH---KDKG-LSVSLSVKNTGVSVAQAQVLAOLYVPLQAAKINPVKLEKFAVE 766
 Db 565 SDIKISSYISDKKAFYVMTYNTSGSRGDEEIIQIYIEPLTPYI-RIKELKGFERYE 623
 QY 767 LQPGETKAVTIEQEKYVAAYFDEERDQVCEKGYEVIVSDSA 811
 Db 624 LKAGESKRVFRLDSSAFYYSKIDHD-WLSESGYNNILIGKSSA 667

RESULT 5

ID 008331 PRELIMINARY: PRY: 754 AA.
 AC 008331;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Beta-glucosidase (EC 3.2.1.21).
 GN BGLZ.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxId=1510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCBI 11754;
 RA Zverlov V., Adelsberger H., Bronnenmeier K., Staudenbauer W.;
 RT "Nucleotide sequence of the bglZ gene from Clostridium stercorarium.";
 RL Submitted (Apr-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCBI 11754;
 RA Bronnenmeier K., Staudenbauer W.;
 RT "Purification and properties of an extracellular beta-glucosidase from
 the cellulolytic thermophile Clostridium stercorarium.";
 RL Appl. Microbiol. Biotechnol. 28:380-386(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCBI 11754;
 RA Schwarz W., Jauris S., Koubka M., Bronnenmeier K., Staudenbauer W.;
 RT "Cloning and expression of Clostridium stercorarium cellulase genes in
 Escherichia coli.";
 RL Biotechnol. Lett. 11:461-466(1989).
 DR EMBL: 294045; CAB08072.1; -
 DR InterPro: IPR002772; GH_3C.
 DR InterPro: IPR001764; GH_3N.
 DR Pfam: Pf00933; Glyco_hydro_3; 1.

DR Pfam: Pf01915; Glyco_hydro_3-C; 1.
 DR PRINTS: PR00133; GLHYDRLASE3.
 DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3; 1.
 KW Glycosidase; Hydrolase.
 SO SEQUENCE 754 AA; 85207 MW; 12D6CEAB28C443EB CRC64;

Query Match 24.98; Score 1095.5; DB 2; Length 754;
 Best Local Similarity 30.34; Pred. No. 1.le-64;
 Matches 258; Conservative 137; Mismatches 247; Indels 209; Gaps 15;

QY 5 DVEALIKKTLAEVVDLAGIDFWHTKALPKHGVPYSLRFDGPNQGV-----GTRK 55
 Db 4 DIKIIISOMTLEERASLCSALDAWTLKGERLGIPSTVMSDGPRLKRRDPDPEK 62
 QY 56 FNGVPACFPQGTSLGTFNQTLLFEAGKMKGEAKIASHAVILPTINNORSPPLGRGF 115
 Db 63 -TVPACFPPLAVGLASWNNRELEVEXKGPALGECQAGIAVLGPETNKLKPLSGRNF 121
 QY 116 ESTGEDPPLAGLAAALIRGIQSTGVQATTIKHPLCNDQEDRRMAYOSIVTERALREIYAL 175
 Db 122 EYFSEDEYLSSEMARSHIKGVOSRGVGTSLKHFANNOEHRMSVDAVIDERTLREIYLA 181
 QY 176 PQQLAVRDSQGFAMFAYNGINGVSCSENPRIYLDGMLRKEMGMDGLMSDWGTYSSTEA 235
 Db 182 SFEQAVKAKAKWTIMCSYINRNGEYASENKFLLTDLVRNMGEGIVSDGAVNERVK 241
 QY 236 VVAGLDLEMPGPPRFREGTELTGFVNSNGKPTLHVIDQAREVLQFVKCAASGVTENGPET 295
 Db 242 LEALDLEMPSSFGIGOKIYEAIVKKELEBEVLDRIVERIILNIFK----- 288
 QY 296 TVNPTPEAA-----LKKVNGEIVLLKNENNVLPLSKKKKTLIVGNPAKQATYHG 347
 Db 289 AYDNKRKNAGYDRANHKLAREARECMVLKKNEDKILPLKOGTLAVIEEFAPKRRYOG 348
 QY 348 GGSALAAIYAAVFPFDLSQLETPPSYTVGATTVTPILGECCLTPDAGPMRMVRVNE 407
 Db 349 GGS----- 352
 QY 408 PPGTRNQHIDELFTKTDMLVDYHPRADTWYADMEGYTADEDCYELGLVVCSTAK 467
 Db 353 -----HVNPTIMDSPYEIKKSAQNNMDVLYAQYII----- 384
 QY 468 KAYVDQLVVDNATKQVPGDAFPGSATRETRERINLVKGNTRYKFIERSAPRYTLKGD 527
 Db 385 -----EKDEP----- 389
 QY 528 IVPGHSLRVGGCKVIDDOAEIEKSVALAKEHDVILICAGLNADMETEGADRASMKLP 587
 Db 390 -----DEKLEBAKOTALKADVAAYFAGLPENHYRCEGDRTHMMPSS 432
 QY 588 LDOLADVAAANPTVVVMQGTPEBFWLADATPAVIOAMYGNETGNSIADVYFGDYNP 647
 Db 433 HCTLIEEVAETVNVVYVLCNGSPVEMPWIDKKGLLEALIGQAMGRLP-IFCSETP 490
 QY 648 -SGKLSLSPFKRLQDNPAFLNFTAGRTLYGEDVYVGRYRYEFADKDVNFPFGHLSYTT 706
 Db 491 IPKLAETFPKQLSDNPSYLNFR-ERDRVREYRGIFVGRYRYKKNMEPLFPGYGLSYT 549
 QY 707 TFAFSNLSVSHKD-----GKLSVSLSVKNTGVSVAQAQVLAOLYVPLQAAKINPVKLEK 762
 Db 550 TFEYGDILKIRKISIDNEVTYVSVKAKNKGMDAGKEIVLYLVDIETSVIDR--RTEPE 607
 QY 763 AKYELQGETKAVTIEQEKYVAAYFDEERDQVCEKGY-----EVIVSDSAK 813
 Db 608 EKVELDQGEKTY-VFELDKRAFAIYHRYKD-WHETGEFRDEIGNSSDIYLLKDKIFVK 665
 QY 814 DGVALRGKFTV 824
 Db 666 STYTIKKRMTV 676

RESULT 6

Q60038

ID 060038 PRELIMINARY: PRT: 720 AA.
 AC 060038;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Beta-glucosidase (EC 3.2.1.21).
 GN BglB.
 OS Thermotoga neapolitana.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NCBI_TaxID=2337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22706-MC24;
 RX MEDLINE=96048478; PubMed=9387232;
 RA Zverlov V., Volkov I., Vellikodvorskaya T., Schwarz M.;
 RT "Thermotoga neapolitana bglB gene, upstream of lamA, encodes a highly
 RT thermostable beta-glucosidase that is a laminaribiose.";
 RL Microbiol. 143:3537-3542(1997).
 DR EMBL: 277856; CAB01407.1;
 DR InterPro: IPR002772; GH_3C.
 DR InterPro: IPR001764; GH_3M.
 DR Pfam: PF00933; Glyco_hydro_3; 1.
 DR Pfam: PF01915; Glyco_hydro_3_C; 1.
 DR PRINTS: PR00133; GLHYDRLASE3.
 DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3; 1.
 KW Glycosidase; Hydrolase.
 SQ SEQUENCE 720 AA; 81370 MW; 4248AD2C75A28BF CRC64;

Query Match 22.4%; Score 982; DB 2; Length 720;
 Best Local Similarity 30.8%; Pred. No. 3.8e-57;
 Matches 265; Conservative 131; Mismatches 275; Indels 190; Gaps 26;

6 VEAILKLTLLAKVDLAGIDFW-----HTKALPKHGVSRLTDPNG 49
 4 VNEILLQLLEKSEKSCG--GWTSGVYVWKSCHGRCGHEHPVPRGLPFLVADBPAG 61
 50 VR-----GTFENGVPACPCGTSIGSTFNQTLLEAKMKGEIAKSAVILGPTIM 105
 62 LRINPTRENDENTYTTAPVPEIMLASTWNRLEEVGKMGEEVREYGVVLLGPAANI 121
 106 QRSPLGGRGPESTIGEDPFLAGIAAALINGIOSTGVQATIKHFLCNDQEDRRMAYOSIVT 165
 122 HRNPLGCRNFEYSEDPVLSGEASAFVQSGVACIKHFVANNQETNMVVDITVI 181
 166 ERALREIYALPFOIAVRDSOPGAFMTAVNGINGVSCSENPXTLDMLKREMGDGLMSD 225
 182 ERALREIYALRGFEIYAVKSKPMSVMSAYNKLNGKTCSONEMILKYLKREMGDFEYVSD 241
 226 WYGYSTTEAVVAGLDLEMPGP-----REFGETLKEVNSNGKPEHIVIDQAREVLO 278
 242 WYAGDNVEQLKAGNDLIMPGKAYOVTERDEIEEIMALKKEGKLEEVLDCEVANIILK 301
 279 FVKKCAASGVTEGPP--ETTANNPTET--AALLKRVNEGIVILKNNVLLPKSKKK 332
 302 -----VLVNAPESEKNRYRSNKPDLKHAHVAAYEAGEGVALLKNE--EALPLSENSK 351
 333 TLIVGPNNAKQTYHGGSAALRAVYAVTPPDGLSKOLETPPSYTVGATYTPVILGEQCL 392
 352 TALEGTGQIEIKGTGSGDHPKXATISILEGKER----- 387
 388 -----GLN---FDEELAKIYEDYIKR-----MRETEEKPR-RDSW----- 419
 453 ECTETELGVVCGTAKAVVDQVLVADNATKQVPGDAFFGSATREETS--RINLVKNTYK 510
 420 -----GTTI---KPLISENPLSEKEVHKL-----AKKNDVAVIYISRLSGEGYD 460
 511 FKIEFGSAPTYLKGDTIVPGHSLRVGGCKVIDDQAEIKESVLAKEHQVITICAGLNA 570
 461 RK-----PVKGDFTLS-----DDETDLITVTS--REFHQ----- 488
 571 DMETEGADPASKMLPGVLDOLADVAANPNTVVVMGTGP--EEMPMLDAPPAVIOAMYG 629

DB 489 -----GKVIYLLNGSPVEYVSMRDLVDGLLWMOA 520
 630 GNETGNSIADVVDYDNPNGSKLSLSPKRLQDNPAL---NRTAGRTIGEDYVYGR 686
 521 GOETGRIVADVLGRINLNGSKLPTFPDRDYSDVSPWTFPEGRKDPKQKVEDIYVGR 580
 687 YTEPADKDVNFPFGHGLSYTTFAPSNLSYSHKDKLSVLSYKNTGSPVGAOVLVYKP 746
 581 YDFTEGPAVEYEGYGLSYTTFEYSDLANSFDEGLTRVGYRIENCTGRAGKREYOVYIK- 639
 747 LQAKINRPVYKELGFAKVE--LQPGTAVTIEOEKYAAAFDEERDQWCEKGYEVI 805
 640 APKGIKDPQELKAFKTRILNPGSEEVLEIPYVDLASFEGEE---WVVEAGEYVR 696
 806 VSDSSAKQDVALGKFTVGE 826
 697 VGASSR--NIKLGTFSSVE 714

RESULT 7

ID 09WKN2 PRELIMINARY: PRT: 721 AA.

AC 09WKN2;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Beta-glucosidase.
 GN TM0025.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MS8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima.*;
 RL Nature 399:323-329(1999).
 DR EMBL: AE001690; AAD35119.1;
 DR TIGR: TM0025;
 DR InterPro: IPR002772; GH_3C.
 DR InterPro: IPR001764; GH_3N.
 DR Pfam: PF00933; Glyco_hydro_3; 1.
 DR Pfam: PF01915; Glyco_hydro_3_C; 1.
 DR PRINTS: PR00133; GLHYDRLASE3.
 DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3; 1.
 KW Complete proteome.
 SQ SEQUENCE 721 AA; 81073 MW; 31DB992D3F4CC963 CRC64;

Query Match 21.8%; Score 959; DB 16; Length 721;
 Best Local Similarity 30.4%; Pred. No. 1.3e-95;
 Matches 262; Conservative 130; Mismatches 280; Indels 190; Gaps 26;

6 VEAILKLTLLAKVDLAGIDF-----WTKALPKHGVSRLTDPNGV 50
 4 IDEILLQLTEKVKLVVGVGLPGLFNGPSRVAAGATGTHPPVPRIGIPAFVLADBPAGL 63
 51 R-----GTFENGVPACPCGTSIGSTFNQTLLEAKMKGEIAKSAVILGPTIMNQ 106
 64 RINPTRENDENTYTTAPVPEIMLASTWNRDLLEEVGKMGEEVREYGVVLLGPAANI 123
 107 RSLPLGGRGPESTIGEDPFLAGIAAALINGIOSTGVQATIKHFLCNDQEDRRMAYOSIVTE 166
 124 RNPPLGCRNFEYSEDPVLSGEASAFVQSGVACIKHFVANNQETNMVVDITVSE 183
 167 RALREIYALPFOIAVRDSOPGAFMTAVNGINGVSCSENPXTLDMLKREMGDGLMSD 226

```

Db 184 RALRELYLKGFBIAYKKAPMTVMASANKLNGKYSQENMLTKKVLREMGDFQVSW 243
OY 227 YGTYSTTEAVVAGLDLEMPGP-----RFRGELKFNNSNGPFIHVLDORAEVLQ 279
Db 244 YAGDNPVEGLKAGNMDIMFGKAYQVNTREKDEIEIMALEKGLSEEVLEDECRNLK 302
OY 280 VAKCAASVTEGPE---TYVNNPPT---AALLKRVNGEIVLKKNNVPLSKKKKT 333
Db 303 -----VLVNAPEFKYRSNKPDLSEHAEVAYEAGAGSVLLEN-NGVLPFDEHTY 353
OY 334 LIVGNPAQATYHGGGSAALRAYAVTPFGLSKQLETPTPSYVAVTVPILGEOCLT 393
Db 354 AVFGIGQLETKGSGSDTHPRITISLEGIK----- 387
OY 384 PDGAPGMFVNEPPEGTPNRQIHDELFTKTDMHLVDYHPRKADPTWYADMEGTYAD 453
Db 388 -----RNMKFEDELASTYEYIK-----MRETEEKKPR-TDSW----- 420
OY 454 DCTYELGVYCGTAKAVVDQVLDVNDATKQVPGDAFFGSATREBTG--RINLVKNTYKF 511
Db 421 -----GTVI-----KPKLPENFLSEKEIK-----AAKKNDVAVVVISRISGEGYDR 462
OY 512 KIEGSAPTYTLKGDITVPGHSLRVGGCKVDDOAEIKSVYALKE-HDOYITACGLNA 570
Db 463 K-----PVKGDFTLS-----DDELEIKTY--SKEFHDO----- 489
OY 571 DWETEGADRASKMLPGVLDLADVAANPNVVMQGTPEEM--PWLDAFPAVIOAWYG 629
Db 490 -----GKVVVILLINISPIEVAWSMDLVGILLVMOA 521
OY 630 GNETGNSIADVVEGDYNPSSGKLSLSPFKRLQDNPAFL--NRTAGRTVLGEDYTYGR 686
Db 522 GQEMGRIVADVGLKINPSGKLTPTPKDYSDVPSPTPEGPKDNPQVYVEEDYTYGR 581
OY 687 YTEFAKQDNVPRGHSLSTTFPAFSLVSHKGLSVLSVKNKTSVPAQVADLYKP 746
Db 582 YTFVFEVPAVEFGYGLSTYKEFYKLDIAIDGELRLRYSTITNGDRAGEVSGYIK 640
OY 747 LQAAKINRPVKELKGAKE--LQPEKTAVTIEDEKVAAYFDEERQMCVEGDIYVI 805
Db 641 APKGIKIDKPFQELKAFHKIKLNPSESEISLEIPLRLASFSDEK--WVESGEYEV 697
OY 806 VDSSSAAKGVALLRGKFTY-GE 826
Db 698 VGASSR---DIRLRIFLVEGE 716

RESULT 8
O51843 PRELIMINARY: PRT: 828 AA.
AC 051843.
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Cellulobiose.
GN CBA.
OS Cellulomonas blazotae.
OC Actinobacteria; Actinobacteriales; Actinobacteridae.
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NBL TaxID=1709.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 486.
RX MEDLINE=9812737; PubMed=9511746;
RA Wong W.K.R., Ali A., Chan W.K., Ho V., Lee N.T.K.;
RT "The cloning, expression and characterization of a cellulobiose gene
RL encoding a secretory enzyme from Cellulomonas blazotae."
DR EMBL: AF005277; AAC38196.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.

```

```

DR Pfam: PF01915; Glyco_hydro_3_C; 1.
DR PRINTS: PR00133; GLHYDLASE3.
DR PROSITE: PS00775; GLYCOSYL HYDROL_E3; 1.
SQ SEQUENCE 828 AA; 86180 MW; 3E8BDE46922030D CRC64;

Query Match
Best Local Similarity 31.2%; Pred. No. 1.8e-55;
Matches 262; Conservative 130; Mismatches 355; Indels 93; Gaps 17;

OY 4 IDVEALIKTLAKVLDLGLIDFEMHTKALPKGVPSLRTDPDNPNGTKFPNGVPAAC 63
Db 7 LDPALVAVSLPELEKVKALLGATRAFTLAPESIGLEGRVLSDEPTVRQLKFSGGRVAL 66
OY 64 FPCGTSIGSTFNQTLLEBAKMMGKEALAKSAHVILGPTINMORSPBLGRGSEIGDEPF 123
Db 67 FPNATLLASAMSEESTTEVGRLLAEELAQOHVIVGLPTINLHRSVGLREASISDL 126
OY 124 LAGLAALIRGIQSTGVQATIKHFLCNOEDRRMAYOSIVTERALREYALPQIAYRD 183
Db 127 LTGRLLAAVYKGLQDVGAGCLHLVANSETERNTMNSVVPATLRELYLPFEIAYDE 186
OY 184 SQGAFMTAVNGINGVSCSENPYLDMLRKEMGMDGLMSDYGYSSTEAHVAGLDLE 243
Db 187 SDPWSVAAVNDVNGVPATEHHVYVNEVLKGEKGTGLVMSMFATRTAAPAAGGLDV 246
OY 244 MPGPFRREGTLKFNNSNGKPFIVHDORAREVLQVKKCAASVTENGPETTYVNNPET 303
Db 247 MPGDGPMGDALVAARVSELDSEVVDHLRLVLAARVGLDRYPDDL--PAPDS 304
OY 304 AA---LIRKVNNEGIVLKNENNVLPLSKKKTLLVGNPAQATYHGGGSAALRAYAVT 360
Db 305 AVREQLTFLAAAGMVLNADDTLEPLARGTRVALVGNHLETTIDMGGSATVPPQVS 364
OY 361 PFGLSKQLETPTPSYVAGYTTVPILBQ--CLTPD--GAGGR-----WRVFENPPG 410
Db 365 VAEGLTALGDADVVDGVEVTRPARPGFVDPDGRGRLHLTLAAGTVLDRHD 424
OY 411 TPNRQIHDELFTKTDMHLVDYHPRK-----ADTYADMEGT 448
Db 425 APS-----LVMMGFDDPFQAAVARFARRVAGALEVGAIGVGRMOTVAGGT 473
OY 449 YTADEDCYELGVCGTAKAYDDQVLDVNDATKQ-----VPGDAFFGSATREBTGR 500
Db 474 -----ELAMTLATSGTGAEMELAPRTIQVHVHGSADVADATVVLRSSFRSVT-- 522
OY 501 INLVKNTYKFKIEGSAPTYTLKGDITVPGHSLRVGGCKVDDOAE---IKSVALK 557
Db 523 -----VGDAADPCTDAGAAAEPLAGYGLGLVARRPAPEAEDDVITRAAAAA 568
OY 558 EHDYITACGLNANMETEGADRASKMLPGVLDLADVAANPNVVMQGTPEEMPW 617
Db 569 QADAVVAVVVGLTBEETESYDKSTIALPGADALVRAVAAAARTVYVVAATPVLMPL 628
OY 618 DATPAVIOAWYGNMTGNSIADVVEGDYNPSSGKLSLSPFKRLQDNPAFLNFTAGRTLY 677
Db 629 DDVAVVIMAGLPQEGGHAVAAALIGDQPTGRVLTTPPAADGAAPAW--SVYPVGDLEY 687
OY 678 GEDVYVGRYEEADK--DVNPFEGHGLSTTFPAFSLVSHKGLSVLSVKNKTSVPA 735
Db 688 TEGFRFVGRGH--WADRAPAPAFMIGHLGATWEYADATLDDGAPANTVYVTNGART 746
OY 736 GAQVAVLYKPLQAAKINRPVKELKGAKEVLEQPGTAKVTIEDEKVAAYFDEERQW 795
Db 747 SREVOVYITPE---ASSDEVR-LVGMADATVDAGASARVVTADAR--MRRWDEAAGW 801

RESULT 9
O87852 PRELIMINARY: PRT: 859 AA.
ID 087852.
AC 087852.
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

```

DE Putative beta-glucosidase.
GN SC06597 OR SC06.18.
OC Streptomyces coelicolor.
OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RX Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares S., Squares S., Taylor K.,
RA Warren T., Wietorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
RC EMBL; AL031013; CAI19790.1; -
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3; 1.
DR PRINTS: PR00133; GLHYDRLASE3.
DR SEQUENCE 859 AA; 89455 MW; BD7EAFEA4FDD59F CRC64;
SO
Query Match 21.8%; Score 957; DB 16; Length 859;
Best Local Similarity 29.7%; Pred. No. 2,4e-55;
Matches 266; Conservative 131; Mismatches 354; Indels 146; Gaps 22;

DB DSPGAGADGAGAPSSGAGELPERGPAHGAKEGSPRRRAGGGRALARRAAGAVLL 361
QY KNNENVALSKKK--KTLIVGNMAKQATYHGGGSAALRAYAVTPEDLSKQL----- 369
DB ANK-DVLPDPHGLCTAVVIGHNAKRTTGGGSGAGVPRREVSVDGIRAEFGRRAYV 420
QY ETTPSYTGAVTTPPIIGECQTP-DGAPGMRWFVN-----EPPGT 411
DB HYPGRPDG---PAPLPDPDCTDPRSGLPVLLMLDADRELAYERRRGRLLEPLV 477
QY 412 ENRQHIDELFTKTDMLVDY---YHPKADTQYADMGTTTADBDCTYEGLIYVCTAK 468
DB PG-----AHVEINARLCPRTGSGW-----SLGVAGFGMS 508
QY 469 AYVDQIVY-----DNATKOYGAFFGSAATREETGRINLYKNTYKIEFGSAP 519
DB 509 LTTDGRTILEGDFPPTSTDDPAVMHVPRAQYTA-----DLTAGR-----D 549
QY 520 TYTLKADTIVPGHG--SLRVGCKVDDQAEIEKSVLAKHEHVOYICAGLADMETEGA 577
DB 550 TLLVARRELAPGTGRATVLAAPAPADVTASLAEAVRAAGADAAYVVVGTTEHGESGY 609
QY 578 DRASKLPGLVDQLADVAANPNTVVMQGTTPREEMDLATPAVIAWGNETGNSI 637
DB 610 DRTDIALGATODALVRAAANAAPRTVAAVNVSGLPELPRDQAGAVLLAMPFGDGGGL 669
QY 638 ADVFEGDYNPSGKLSLSPFKRLQDNPAFLNFTTEAGRTLYGEDVYVGYRYEFAKDYNF 697
DB 670 ADVLRHAPGGRRLPTTAVLADAPV-TRTRPDGRLDYDGLHNGHMLRHRTAY 728
QY 698 PEGHGLSTYTFAPSNLS---VSHKDGKLSVSVKNTGSPVQAQVQLYV-KPIQAAIN 753
DB 729 WEGHGIGYTTMYEELTLPVPTRAGDGLTVRVNRVTGARAREROVVOYVLAIP-ASALD 786
QY 754 RPVKELKGANVYELQGEKATVIEQEKVYVAFDEEDDOMCEVGEDEVIVSDSS 810
DB 787 RPARWLAGYTAARARGETVATVTRYPAR-ALRHMSVAEHAKRTAGPCRVLAGNSA 842

RESULT 10
ID 070021 PRELIMINARY; PRT; 808 AA.
AC 070021;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Beta-glucosidase.
GN ERYBI.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
OC Saccharopolyspora.
NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL2338;
RX MEDLINE=98273631; PubMed=9613575;
RA Gaiser S., Boehm G.A., Donlith M., Raynal M.C., Dhillon N.,
RA Cortes J., Leadlay P.F.;
RT "Analysis of Erybi and ErybIII from the erythromycin biosynthetic gene
RT cluster in Saccharopolyspora erythraea.";
RL Mol. Gen. Genet. 258:78-88(1998).
DR EMBL; Y14327; CAA74702.1; -
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3; 1.
DR PRINTS: PR00133; GLHYDRLASE3.
SO SEQUENCE 808 AA; 86677 MW; OC619016CC00A751 CRC64;
Query Match 21.7%; Score 954; DB 2; Length 808;
Best Local Similarity 30.8%; Pred. No. 3,4e-55;

Matches 260; Conservative 128; Mismatches 334; Indels 122; Gaps 21;

```

OY 6 VEALTKLTLAEKVDLAGIDFMH-----TKALPKGVSLRTPDQPNVGRGK 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 VDLIQLQLLDEKLSFV-----YWDYNEKPLAKMLPGVPLIGTQIRGDSGPAGVTHIQ 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 55 FENGVPACFGCTSLGTFNQTLLEBAKMMKKEALSAHVILGPTINMORSPILGGRG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 -----PAIMPAAPVALASAFDDRLAHEYGTVLGREGRAFEDQILLGPVNNIRVPOAGN 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 FESIGDPPPLAGIAGAAALRGISTQSVQATIHFLCNDQEDRRMAYQSLVERALREIYA 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 FEFESDPLVLTARTAAQIRGHISOGMLMSAHYANTQEDRTIDVDQRTIRETEL 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 175 LPFOIARVDSOGAFMTAVNGVSCSENPXYLQGMRLKREMGDGLMSDYGTSYTE 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 PGFEAAVA-AGATSVWCAIPKYNNGHACHRRLLLEIKQMGFGVWMSDWTATHA-TE 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 235 AVYAGLDLEMPGPR-----FRGTLTKFVNSNGKPFTHVIDORAREVLOFYKCAA 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 DLVAGIDQMGVEVEDGSLFRGKYLEALKKAIREGRIPESSALDASVRIITQFEREGL 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 286 SGVTENGPEPTVNNTPETALLRKVNGNEVILKNENNVLP--SKKKTLIVGNAAQA 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 --LDETKPPRPBRDVAAGGTRIAQEVASEGAVILRNEGGVLLDPAAGDIAVIGSAQP 383
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 344 TVHGGSAALRAYAVTPEDLSKOLET--PSSYVGAATYVPPILGQCLTPDPAQPMR 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 KYTGAGSSYVEBDPANAPLDTITQRVGSGRGVSGVELGAPL-PTALQAPNVPG-- 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 402 WRFVNEPPTPKOHIDELFFTKTMHLVDYTHPKADTWYADMGTYTADECTYEGL 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 -----EVTPE-----PPSGGVYI--DGRILKVDAGLYRTAA 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 462 VVCGTAKAVVDQLVVDNATKQVPDGAFFGSATREETGRINLVKNTYKFIKTEG--SA 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 RIDG-----GNGSLQIDGAPIG-----VGDV-----FGPLTSV 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 519 PYYTLKGTIYFGHSLRVGCKVID-----QAEIKSVALAKENHOVITICAGLN 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 PWWLTKGHEITOMTGAAGVVGSLDLDVLTWYPGHAQREFPAVAERARDSDVAVVFA-- 554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 570 ADMETEGADRSMKLPGLVDLIANDVAANNTVVMOTGPBEKPMIDATPAVIOANYG 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 YDDGELTADRTSLPLGQDKLIDIVASVNPNTVYVLTGSSVTMPMDKTRAVLDMWYR 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 630 GNETGNSIADVYFGDYNFSGKLSLSPKRLQDNPAF--LNFRTBAGRTLIGEDVYGYR 686
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 GQAGAEMATTALLFGDAEFGRLGTFFPYSGERTPYGDPARFPGVDGVHSEGIFFSGYR 674
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 687 YTEPADKVNPPFGHSLYTPFANSLSVSHKDGKLSVSVKNTGVSVPQAQVQLVYKP 746
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 WYDRGCVPLPFPFGHSLYTPFERTRDGLDVTIVRNTGORGSDVYQVYVLP 734
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 747 LQAAKINPVKELGFAVELQPGETKAVTTEBOEKYAAAFEDDERQWCEKEDYEVY 806
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 SPQVPLDAPAKQLAGYKVELAPGETKRVYNAER-ALQHMWDAAGWKLGGKRAVEI 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 807 SDS 810
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 794 GSS 797
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

09A7G7 PRELIMINARY; PRT; 762 AA.
 AC 09A7G7;
 DT 01-JUN-2001 (TIREMBLrel. 17, Created)
 DT 01-JUN-2001 (TIREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TIREMBLrel. 20, Last annotation update)
 DE Beta-D-glucosidase.
 GN CC1756.
 OS Caulobacter crescentus.

```

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter
OC NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX Medline-2173698; Pubmed-11259647;
RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eelsen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Emolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005849; AAK23732.1; -
DR TIGR; CC1756; -
DR InterPro; IPR002772; GH_3C.
DR InterPro; IPR001764; GH_3N.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF009933; Glyco_hydro_3_1.
DR Pfam; PF01915; Glyco_hydro_3_C; 1.
DR PRINTS; PR00133; GLYDRLASE3.
DR PROSITE; PS00436; PEROXIDASE_2; UNKNOWN_1.
KW Complete proteome.
SQ
SEQUENCE 762 AA; 81011 MW; CALC76B0D7E0ZDB CRC64;

```

Query Match 21.6%; Score 946.5; DB 16; Length 762;
 Best Local Similarity 30.3%; Pred. No. 9, 9e-55;
 Matches 257; Conservative 115; Mismatches 273; Indels 199; Gaps 20;

```

OY 4 IDVEALTKLTLAEKVDLAGI-----DFWHTKRLPKHGVSLRFTGPN 49
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 VRADALVKQMTLDEKITYLHGLFRPMAKRAPDTMISAGVPEVPLNPTLRPSASIG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 50 VRG-TKFNQVPAACFGCTSLGTFNQTLLEBAKMMKKEALSAHVILGPTINMORS 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 VANQVEQRKGDVATLPSGLALSTFEPKLAFGAMIGAEARAKFPNVLLAGVNLTRD 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 109 PLGGREFSIEGDEPFLAGLAALINGISTQVQATIKHLCNDQDRRMVQSIYTERA 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 PMAGRNFEYIGEDPLAAGENVAQIKGVOSNKTIVSTIKHMLAOTGRIVYDAQIDED 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 169 LREIYALPQIAVRDSQGAFTAVNGVSCSENPXYLQGMRLKREMGDGLMSDYG 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 LRSDDLAFQIALEKSNPASMCAVKNVGDACENDPLNKLKDKMNYPGVWSDMKA 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 229 TYSTTEAVAGLDLEMPGPR--FRGTLKFNVNSNGKPFTHVIDORAREVLOFYKCA 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 VHSIVKAAALGLD-QOSQGLDTQIFFGEDLKAAVAKGEVSQARVDVNRILHGV--I 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 285 ASGVTEGPEPTVNN--NTPETALLRKVNGNEVILKNENNVLPSSK-KTTLIVGNPK 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 SSGLMNPTTSQAPLDYAHAAVQTVABERGVSILKNRGDLPLAKSKKTVLLGAAHD 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 342 QATYHGGSAALRAYAVTPEDLSKOLET--PSSYVGAATYVPPILGQCLTPDPAQPMR 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 VGVISGGSSQVR-----SVGARVPEIPLKGBAAS----- 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 402 WRFVNEPPTPKOHIDELFFTKTDMHLVDYTHPKADTWYADMGTYTADECTYEGL 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 -----EVTPE-----PPSGGVYI--DGRILKVDAGLYRTAA 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 462 VVCGTAKAVVDQLVVDNATKQVPDGAFFGSATREETGRINLVKNTYKFIKTEGSAPTY 521
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 -----SSPLQ 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 522 TLKGDITVPGHSLRVGCKVIDQAEIEKSVLAKENHOVITICAGLNADMETEGADRAS 581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 AIKA-----ANPSAEV--TYVDGKDPV-AAAAAAKADADVAIVPAW--HMQTEADAPS 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 582 MKLPGLVDLIADVAANANTVVMOTGPBEKPMIDATPAVIOANYGNETGNSIADV 641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db      503 IALPENODALIEAVSANKNAVVLEETGCVLMPWLCKGAVLQWPGQGGQALAIL 562
Qy      642 EGDYVNSGKLSLSEPKRLQDN-----AFLNFTREAGR-----TLXGEDVY 682
Db      563 FGEVNFSGRIAMTFPSEDOAPRASPAPRAEQALIDARAGOKKAPIKGFVRYVEGMA 622
Qy      683 VGYRYVEFADKDVNFEGHGLSYTTFASFNLVSYSKDKLSVLSYKNTGSPGAQVACL 742
Db      623 VGYWFMQEKRRPLPYPGYGLSYTSPGYKNLKVEDGDG-LKVSFDVTNTGKVAADTDL 681
Qy      743 YVFLQAKINRPYKELKAGAKYELQGETKATIEOEKRYAAYADEERDQMCYEKGY 802
Db      682 VYTSQO-----RKAMLRLAGFOKVLAPGETKRVTLNVEPRILADY-DTAKPGMTIAGTY 736
Qy      803 EVIV 806
Db      737 FLIV 740

```

```

RESULT 12
O9ZGH5 PRELIMINARY; PRT; 809 AA.
ID O9ZGH5
AC O9ZGH5
DT 01-MAY-1999 (TREMBlrel, 10, Created)
DT 01-MAY-1999 (TREMBlrel, 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel, 20, Last annotation update)
DE Beta-glucosidase.
GN Desr.
OS Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=54571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC15439;
RC MEDLINE=98445333; PubMed=9770448;
RA Xue Y., Zhao L., Liu H.W., Sherman D.H.;
RT "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT venezuelae: architecture of metabolic diversity."
RT Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL: AF079762; AAC68679.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro.3; 1.
DR Pfam: PF01913; Glyco_hydro.3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
SQ SEQUENCE 809 AA; 84686 MW; DE15489926BE37A8 CRC64;

```

```

Query Match 21.5%; Score 944.5; DB 2; Length 809;
Best Local Similarity 32.2%; Pred. No. 1.5e-54;
Matches 275; Conservative 112; Mismatches 325; Indels 143; Gaps 26;

```

```

Qy      9 ILKLTIAEKVDLLAGIDFW-----HTKLPKHGVSRLRTDGPNGVGRKTFENG 58
Db      56 LVAMQTLIDKISFV-----FMALDPRONVGYLPVPRIGIDELRAADPNRIR---LVG 107
Qy      59 VPAPCPGSGISLTGNOLLEAGKMGKKAIAKSAHYIIGPTINMORSPGLGGRFESI 118
Db      108 QRTALPAPVALASTFDYDADSYGVNMRDGRALNODMVLGPMANNIRVHGSGNYTF 167
Qy      119 GEDPFLAGLAALIRIGISTGVATIKHFLCNDQEDRRMAYOSTIVERALREIYALFQ 178
Db      168 SEDPVSRSRAVAOIKIGOGAGIMTAKHFAANNOENNFSVNAVNDQOTLREIEFPFE 227
Qy      179 IAVRDSOPGA-TWTAYNGINGVSCENSKYIDGMIRKMGMDGLIMSVWYTSSTAY 236
Db      228 AS---SKAGASPMCAVANGKPSGDNELNNVLRITQMGQGVMSMLAT-FCIDAI 283
Qy      237 VAGIDLFM-----PGPP-RFGETLKFVNSNK-PFIHYIDQARAVLQFYVK 282
Db      284 TKGIDQMGVGLPGDVPKSGPAPKPFGEALKTAVLNGTVEAATVRSABRIYQOMEKF 343

```

```

Qy      283 CAASGVTENGPETTNNTEETALLRKVGNEGIVILKNNENNVPLS--KKKTLIVGPA 340
Db      344 GILLATPAPRPE---RDKAGAGVSRKVAENGAIVLLRNKGALPLIAGAGKIAVIGPTA 400
Qy      341 KQATYHGGGSAALRAVYATYTPFDGLSKOLEPSPSYVGAATVVPPLIGOCITPGAGGM 400
Db      401 VDPKVTGLGSAHVAPPDAAAPLDITIKAR-----AGAGATVYETGETGTGTGIPAGN 452
Qy      401 RWRVNEPPTPNRQHIDELFTKTDMLVDYVHPRAADTWYADMEGYTAEDCTYELG 460
Db      453 LSPAFNQG-----HOLE--PGKAGALY---DGLTLPADDEYRLA 487
Qy      461 LVWGTAKAVYDDQLVVNDATKQVPGDAFFGSAATRETRINLVKNGTYKFIIEGSAVT 520
Db      488 VRATG---GYATVGL---GSHTEAGQVYKYS---SPLKLTG-TKLL-----527
Qy      521 YTLKGDITVPHGSLRYGGCKVIDDOAEIKSVLAKKEHDVITICAGLADMETEGADRA 580
Db      528 -TISGFAMASATPLSLLEIGWVTPPAADATIKAKVESAKKATAVFA---YDGTGEGVDR 583
Qy      581 SMKLPVLDOLIALDVAANPNTVVVVMOTGTPPEMPWIDATPAVIAQWYGNELGNSIADY 640
Db      584 NLSLEPGTDKLSAVADANPNTIVLVNTGSSVLMPLSKTRAVLDMWYPCZAGAEATAAL 643
Qy      641 VEGDYNPSGKLSLSEPKRLQDNPAFLNFTREAG-----RTLYGEDVYGYRYEE 690
Db      644 LYGDVNPSSKLTQSP-----PAENQHAVAGDPTSYPCVDMQOITYREGIHYGYMFK 696
Qy      691 ADKDVNPFEGHGLSYTTFASFNLVSVH-KDGKLSVLSVKNRTGSPVGAQVAYVYKPLQA 749
Db      697 EVKRLPFFPGHGLSTSTQCAPTYVRTSGGLVATYVTRNSGRAGQEVQAYLGAAPN 756
Qy      750 AKINRPVELKFAKVELQGETKAVTIEOEKRYAAYFDEERQMCVEKGDVYVSDS 809
Db      757 VTAQAKKKLVGYTVKVSIAAGEAKTVTVN-----DRROJOTGS 795
Qy      810 SAKAGVALRGFTV 824
Db      796 SSAD---LRGSATV 806

```

```

RESULT 13
O9P3Y0 PRELIMINARY; PRT; 781 AA.
ID O9P3Y0
AC O9P3Y0
DT 01-MAR-2001 (TREMBlrel, 16, Created)
DT 01-MAR-2001 (TREMBlrel, 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel, 20, Last annotation update)
DE Exo-1,4-beta-glucosidase (EC 3.2.1.21).
GN Prevotella albensis.
OS Bacteria; CFB group; Bacteroidetes; Bacteroidales; Prevotellaceae;
OC Prevotella.
OX NCBI_TaxID=77768;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M384;
RA Walker N.D., McMan N.R., Wallace R.J.;
RT "exo-1,4-beta-glucosidase gene from Prevotella albensis."
RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ276021; CAC07184.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro.3; 1.
DR Pfam: PF01915; Glyco_hydro.3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
SQ SEQUENCE 781 AA; 85866 MW; 7F025625B8735CB5 CRC64;

```

```

Query Match 21.4%; Score 941; DB 2; Length 781;
Best Local Similarity 29.6%; Pred. No. 2.4e-54;
Matches 257; Conservative 123; Mismatches 289; Indels 200; Gaps 24;
Qy      5 DVEAIIKLTIAEKVDLLAGI---DF-----WHTKLPKHGVSRLRF 43

```

```

Db 31 NIDVIAAMTLEKKAOLLVGVGHODVSSGTMIGQHSRLVAGAMAGOTAEISRLGIPATV 90
QY 44 TDSPNCR-----GTR--FENGVPACPCGTSIGSTFNOCTLLEAGKMKKEAIKAKSA 95
Db 91 ADGPAGVHINPRTPTGNQF-----ATGPRTICTLASTWINDVLYHVAKAIGNETLEYGI 146
QY 96 HVLIGFTIMQSSPLGGRFESIGEDPLAGIAAALIRGIQSTGVQATIKHFLCNDQED 155
Db 147 DVLIGGMNLHSPLOGRNFEEYSDBPIVYGLIGSAMKGIQSGVSAKFAHANSSES 206
QY 156 RRMVQSIYTERALREIYALPROIAVRDSOPAEWTAANGVINGSCSNRPYIDGMIRKE 215
Db 207 DRRVRDERISQALREIYALKEIYVRSKPTMTLSSTNKNGTYTQSKDNLNIRK 266
QY 216 WGMGLIGMDWYGYSS-----TTEAVVAGLDEMPGPRFRGETLKFVNSKGFPIHYIDQ 271
Db 267 WCYOGIVMTDMIGERADIPVETE--VEAGNDFMFGND--RAKHLYKAKVAGRLDIKIVAR 324
QY 272 RAREVLOEFKCAASGVTEENGPEETVNTPEP-----AALIKRVNEGIVLLKNNVPLPS 328
Db 325 NIKNNLEIYTLK-----TPRTKTKYTYNQPOLKAHAQITTRQASTEGVLLKNDNNVLPVK 378
QY 329 KKKKTLIVGPNKQATYHGGSSAALRAYVAVTPEDGLSKOLETPPSYTVAGATVPPIIG 388
Db 379 NKKKVALRGVNSYDPLSGLSGCNNVPYVDMVHGLQ-----NAGIATP----- 423
QY 389 EGCCLTPDAPGKMRVFNPEPPTNROHIDELFTTKTDMHLVDYHPRKADTYVADMEGT 448
Db 424 -----KQLEIYENYVAKAKLQADKNEMFELDQ--- 454
QY 449 YTADECTYELGLVCGTAKAYVDDOLVVDNATKQVDFAFGSAFPEEIGRLINLYKNT 508
Db 455 -----GQPK--LDEIETQRCVHEVGDA-----DAIITIKR--- 485
QY 509 YKFKIEPSAFTYTLKGDITVPHGSLRVGCGKVIDDOAEIEKSVALAKEHDVITICGL 568
Db 486 -----QAGEGMDRSIEGEFNLTDH-----EKAMISR----- 511
QY 569 NADMETEGADBASMKLPGVLDQLADVAAAN--PNTVYVMOGTPEEMPLATPAVIOA 626
Db 512 -----VSDVFHANNKPVYIILNNGSVMEETASMRDVAIIVA 548
QY 627 WYGNETGNSIADVYFGDYNPSCGLSLSP-----KRLQDNPAFLNFRTE----- 671
Db 549 WQPEEGEGNSYADVLLIGKANPSGHLSTWPIISADVPSTKNPQQAAYINLSDKLYSNM 608
QY 672 --AGRTLYGEDVYVYGYRYEFADKDVNFPBGHLSITTAFSMLSVSHKDGKLSVLYK 729
Db 609 KGVNVTNHEEDIVGYRYFPTFNKKVAYPPGYGLSYTTEFGKPSVSLNGDKITVVSVK 668
QY 730 NTGSVPGAQVAYOLVYKPLQAAKINRPVKELGFAKVE--LQPGETKAVTIEBQEKVAAVF 788
Db 669 NIGVACKQVAYVYVAPKGA--YKPSCELEKAPAKTKNLKPOSEETIKMTAKRLDAS-F 726
QY 789 DEEDDMQCVKEKDEVIYVSDSSAAKDGA 817
Db 727 DEANSQMKYDAGKYEEFCVGDWVESIKGTA 755

RESULT 14
082C29
AC 082C29 PRELIMINARY; PRT; 793 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative beta-glucosidase (EC 3.2.1.21).
GN EGLB OR YP02803.
OC Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;

RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagsels K., Kariyasek A.V.,
RA Leather S., Koutle S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
EMBL: AJ141153; CAC93037.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3; 1.
DR Hydrolase; Glycosidase; Hypothetical protein; Complete proteome.
KW SEQUENCE 793 AA; 86853 MW; 6C23D603998955 CRC64;

Query Match 21.4%; Score 939.5; DB 16; Length 793;
Best Local Similarity 28.7%; Pred. No. 3.1e-54;
Matches 242; Conservative 125; Mismatches 266; Indels 211; Gaps 16;

QY 6 VEAIIKRLTLEKVDLLAGIDF-----WITKALPRKHGVPSLRFIDGPNCR----- 51
Db 31 IPSYIAAMTLEKKAFAVSGTGAMNAEKVAVAGASTLAIPRLGIPQIVFADGPGVRLGAG 90
QY 52 ---GTRFENGVPACPCGTSIGSTFNOCTLLEAGKMKKEAIKAKSAHVLIGFTIMQSS 108
Db 91 PTGGEKRF-----ATGPFVSAVAMATMDPALIKRVGALIGDEAQQVGVLDILGPAIQNH 146
QY 109 PLGGRFESIGEDPEFLAGIAAALIRGIQSTGVQATIKHFLCNDQEDRRMAYQSIYTERA 168
Db 147 PLNGRNEFYTEDPLNAINATYASTIDQMOTEGVAGALKHFAANNQDRETRQIYNEIISDRA 206
QY 169 LREIYALPFOJAVRDSQPAFMTAYNGINGSCSENPKYLDGLRKEMGMDGLINSDWG 228
Db 207 LHEIYFPGFEYAEKQAPMAVMSYSPSINGTSPSSQNPMLIKVLRQOKMFDGFVMSDWYG 266
QY 229 TYSTTEVAVAGLDEMPG-----PFRGETLKFVNSKGFPIHYIDQAREVL 277
Db 267 VADVPALKGNGNLDNMGGRTPDPSLFLTPNTDPRKVVYLAALKSGELQDQIDENIRIL 326
QY 278 QFVYKCAASGVTEENGPEETVNTPETALLRKVNGSGIYLLKNNVPLI,SKKKKTLIVG 337
Db 327 NVYIK-----THRFKNGDAAMMPEDVDHSAKARQIADSMVLLKNDHATLP,IKAAFPVAAFG 383
QY 338 PNKQATYHGGSSAALRAYVAVTPEDGLSKOLETPPSYTVAGATVPPIIGROCLTPDGA 397
Db 384 QNIDNFEVTVGSGSAEVN-----IDP-----KRLVTLPQGLKKNAGIALINQIQDAL----- 429
QY 398 KGRKWRVFNPEPPTNROHIDELFTTKTDMHLVDYHPRKADTYADMEGTYTADECTY 457
Db 430 -----DEPVNP----- 436
QY 458 ELGLVVGCTAKAYVDDOLVVDNATKQVDFAFGSAFPEEIGRLINLYKNTKRIEFGS 517
Db 437 -----ALTKA----- 441
QY 518 APTYTLKGDITVPHGSLRVGCGKVIDDOAEIEKSVALAKEHDVITICAGLNADMETEGA 577
Db 442 -----AQSDDLALISIGSS-----TEGA 460
QY 578 DRASMKLPGVLDQLADVAA-----NPTVYVMOGTGPEEN--PULDATPAVIAQWYSGNE 632
Db 461 DRYSAMHPDEVAMIKTKYASAFHQOKRVVYLLNIGSPIEMTSEWQYADAILTLWQGEQ 520
QY 633 TGNSTADVYFGDYNPSCGLSLSPKRLQDNPAFLNFRTEAGRTLYGDDYVGVGRYYFAD 692

```

Db 521 AGNAVADILLGKVPNSCLPLTPFKRLDESPSGNPGNAKTVYIGGIYGVYRFPDKR 580
 QY 693 KDVNPEGHLSYTFTEAFNSLVS-----HKDGKLSVLSYKNTGSGVGAOVQOLVYKP 746
 Db 561 IAPMPEFGISYTVNGKISPEKPVFNIDTENSIEVSTVNTSGIDTKEVQLYVHD 640
 QY 747 LQAKINRVEVELKFAKVELQPEKTAVTIEOEKYAAVFEDEPDQWCEKGDYEVIV 806
 Db 641 -NASTLRPEQELAKFAEKVSLSAGEKRYAF-KIDKRAFSTYDDKNNWVLEPOLFTLRI 698
 QY 807 SDSS 810
 Db 699 GRSS 702

RESULT 15
 Q9RH03 PRELIMINARY; PRT; 732 AA.
 ID Q9RH03
 AC Q9RH03;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Beta-glucosidase precursor.
 GN SALA.
 OS Azospirillum irakense.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 OC Azospirillum.
 OX NCBI_TaxID=34011;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KBC1.
 RA Faure D., Desair J., Keljers V., Proost P., Hentissat B.,
 RA Vanderleyden J.;
 RT "The sala and salb genes encode two beta-glucosidases in Azospirillum
 irakense.";
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF090429; AAF21798.1;
 DR InterPro; IPR002772; GH_3C.
 DR InterPro; IPR001764; GH_3N.
 DR Pfam; PF00933; Glyco_hydro_3; 1.
 DR Pfam; PF00933; Glyco_hydro_3_C; 1.
 DR PRINTS; PR00133; GLHYDLASE3.
 KW Signal.
 FT SIGNAL.
 SO SEQUENCE 1 26 POTENTIAL.
 Query Match 20.9%; Score 919.5; DB 2; Length 732;
 Best local similarity 30.7%; Pred. No. 6e-53;
 Matches 260; Conservative 106; Mismatches 275; Indels 205; Gaps 23;

QY 12 KLTLAEKVDLAG-----IDFWHTKALPKHGVPSLRFTDGPNGVR 51
 Db 49 ELTIDERSILHGPMPPLFPFGSPPIPEGPSLVPIYF--PGVPLGIPALKETDASLGV- 104
 QY 52 GTRFEF--GVPAACFGCTSLGSTFNQTLLEAGKIMGKEAIAKSAHVILGPTINQRS 108
 Db 105 -TNPNNVAPGPTATALPGALASTFNPKLISYDGAALAKKAASKGFNVLLAGANLARD 163
 QY 109 PLGGRGFESTIGEDPPLAGLAALIRGIQSTGVQATIKHPLCNDQEDRRMVSIVTERA 168
 Db 164 PRNGNFEYLEDPLLAGIAGESIRGIOSONIISTVKGHESLNGQETNRHMGNSVIDEAA 223
 QY 169 LREIYALPEQIAVDSOGAFMTAYNGINGVSCSENPKYLDGMLRKEMGNDGLMSDYG 228
 Db 224 HRESDLAFQIAIERGDPGSMCAVNLVNGAYSCGNDHLNKLKYLKGDWGYKGMVMSDW-G 282
 QY 229 TYSTTEAVVAGLDEMPGPR---FRGETLKFNVSNKGPFIHVIDQAREVLQFVKCA 284
 Db 283 AVPATDFALKGLD-QQSGQQLDEKIMFQDILKEAAAAGTIPAERLSDMSRRIL---RSMF 338
 QY 285 ASGYTENGPETVNTPTALLKRVNGEIVLLKNNVPLSK-KKTLIVGPNAKQA 343
 Db 339 AAGFDDGKPRPVVDLHAHAIAKOVADGIVLLANDKGLPLAAGSOKIAVIGGFADQG 398

QY 344 TYHGGSAALRAYAVTVPEDGLSKOLETPPSYTVGAVTVPPIIGBOCLTPDQAPGMRR 403
 Db 399 VLSGAGSSQYT-----SVGNPVIPIPGSGML-----AATLRQA 433
 QY 404 VNEPPGTPNROHIDELFTKTDMLVDYHHPKRAADTWYADMGTYTADECTYELGLVY 463
 Db 434 YHNSPLKALKERL-----PNATIRF---NDGRYSA----- 461
 QY 464 CGTAKAYVDQLVYDNATKQVPGDAFFGSAKTRBETGRINLVKNTYKFKIERGSAFTYTL 523
 Db 462 ----- 461
 QY 524 KGDTIYVGHSLAVGGKVIDDQAEIEKSYALAKEHDYVLIICAGLNADMETEGADRASK 583
 Db 462 -----AAALAKQSDIYILFAN---QWMSBGMADYDLK 491
 QY 584 LPGVLDQLADVAANPNVTVVWQGTPEEMWLDATPAVIOAWYSGNETGNSIADVVG 643
 Db 492 LPGQDALIIEVAEAMPNAVIVLQGTGPVLMPKDKVGAIVSAWYSGQKGGELADILVG 551
 QY 644 DYNPCKLSLSEPKRLQ-----DNPAFLNRTBAGRTLYGEDYVGYRYTFRADKVN-- 696
 Db 552 KTNPSGLPSTFPASADQYPHPEVPGMNLPEKQOPDVVEEGSDVGYR--RFAAKGMKPL 609
 QY 697 FPPGHGISTYTFAPNSLVSXKDGK-LYSVLSYKNTGSVPGAQVQOLVYKPILOAKINRP 755
 Db 610 FPPGHGISTYTFAYDKLV--KGETLEVSFOVTNKGKLOGKHAPOIYTLAGANGQILQR- 666
 QY 756 VKELGFAKVELQPGETKAVTIEOEKYAAVFEDEPDQWCEKGDYEVIVSDSSAAKDG 815
 Db 667 ---LIGFEKIDLPKERRRVTIKADRLRL--FDEGQHWRIDGSDYDVVVGSAATM--- 719
 QY 816 VALRGK 821
 Db 720 TVLSGR 725

Search completed: April 26, 2003, 13:09:46
 Job time : 52 secs

